

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

## Relationships between yield and pollen concentrations in Chilean hazelnut orchards

### This is the author's manuscript

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/1737716> since 2020-04-29T13:02:43Z

*Published version:*

DOI:10.1016/j.eja.2020.126036

*Terms of use:*

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)



# Relationships between yield and pollen concentrations in Chilean hazelnut orchards

Lorenzo Ascari <sup>a,\*</sup>, Consolata Siniscalco <sup>a</sup>, Giovanna Palestini <sup>a</sup>, María José Lisperguer <sup>b</sup>, Eloy Suarez Huerta <sup>c</sup>, Tommaso De Gregorio <sup>c</sup>, Simone Bregaglio <sup>d</sup>

<sup>a</sup> Department of Life Sciences and Systems Biology, University of Turin, Viale Mattioli, 25, 10125 Turin, Italy

<sup>b</sup> Fruticola Agrícola S.A., Lote A Higuera 1 La Florida del Alto, Curicó, Chile

<sup>c</sup> Hazelnut Company Division, Ferrero Trading Luxembourg, Rue de Trèves, L 2632 Findel, Luxembourg

<sup>d</sup> CREA - Council for Agricultural Research and Economics, Research Centre for Agriculture and Environment, via di Corticella 133, 40128 Bologna, Italy

## ARTICLE INFO

### Keywords:

Gaussian model  
Logistic model  
Pollen dispersal  
Hazelnut production  
Yield variability

## ABSTRACT

Chilean hazelnut (*Corylus avellana* L.) industry has been recently experiencing a solid development and nut production is consistently increasing. *Corylus avellana* is subjected to strong inter-annual yield fluctuations, which raise the issue of price predictability. In alternate bearing, a high fruit load hinders flower initiation thus leading to a smaller fruits production in the following year. This suggests the use of pollen airborne concentration as a proxy for flowers number to predict hazelnut irregular bearing. We addressed this question using high-resolution yield and pollen data collected on four hazelnut plantations located in Central Chile, and three hazelnut cultivars. Pollen parameters were calculated from modelled and non-modelled pollen seasons, after selecting the best modelling approach. An alternate bearing index was calculated to characterize hazelnut biennial bearing and to select plots with high yield variability. The calculated pollen parameters were then correlated with yield data. Significant correlations were found between pollen parameters and nut yields, with stronger relationship when alternate bearing was higher. The modelled peak of the pollen season presented the highest variability (average CV = 81 %) and its positive correlation with crop production ( $\rho = 0.45$ ) was consistent across sites and cultivars. Our results support the integration of airborne pollen data in hazelnut yield prediction systems, as predictors of alternate bearing and to forecast annual yield.

## 1. Introduction

European hazelnut (*Corylus avellana* L.) is an important economic nut crop worldwide. Following the common trend in nuts, the global hazelnut cultivated area intensively enlarged in the last five years, exceeding 670,000 ha in 2017 (FAOSTAT, 2020). Moreover, hazelnut cultivation recently expanded outside the native range of the species, i.e., Europe and Western Asia, when farmers in the Southern Hemisphere started to plant new hazelnut orchards.

*C. avellana* was firstly exported in South America by European immigrants more than 200 years ago and was recently introduced in Chile, where commercial orchards were established in the 90s and which nowadays is the seventh hazelnut top producing country (Nuts and Dried Fruits Statistical Yearbook, 2020). The hazelnut area in Chile 13,700 ha in 2017 has progressively expanded at an average rate of 2500 ha year<sup>-1</sup> since 2010. The Chilean in-shell hazelnut production has consequently increased from around 1000 tons in 2007 to

more than 18,000 tons in 2017 (FAOSTAT, 2020). As a recently introduced crop in this country, hazelnut is experiencing new agro-environmental conditions in terms of climatic variability and pedologic characteristics, whose impacts on yield variability has not been yet investigated.

Despite the long domestication history and breeding activities (Molnar, 2011; Boccacci et al., 2013), both hazelnut wild accessions and cultivated varieties share a marked irregular bearing, leading to large inter-annual variation of global hazelnut yields (FAOSTAT, 2020). This phenomenon, which is common to other fruit crops, is known as alternate or biennial bearing and consists in a higher fruit production in the ON year which inhibits flowering in the OFF year, hence leading to lower fruit load (Smith and Samach, 2013). Many authors suggest that airborne pollen concentration could be then used as an early indicator of ON and OFF growing seasons, e.g., on holm-oak (García-Mozo et al., 2007), as well as a predictor in yield forecasting systems, e.g. in olive (Galán et al., 2008; Orlandi et al., 2010; Oteros

\* Corresponding author at: Viale Mattioli, 25, 10125 Turin, Italy.  
Email address: [lorenzo.ascari@unito.it](mailto:lorenzo.ascari@unito.it) (L. Ascari)

et al., 2014; Ribeiro et al., 2017), grapevine (Cunha et al., 2016), and walnut (Prentović et al., 2014).

*C. avellana* is a monoecious, wind pollinated species (Germain, 1994). Pollen dispersal is strongly influenced by local climatic conditions and can start as early as middle November in early cultivars flowering in mild climates and last until the beginning of March in colder conditions and late cultivars, in the Northern Hemisphere (Solar and Stampar, 2011; Cristofori et al., 2018). Self-pollen is rejected by the stigmas due to sporophytic self-incompatibility, and compatible cross pollen is required to fertilize the ovules (Heslop-Harrison et al., 1986; Hampson and Azarenko, 1993), even if partial or complete self-compatibility was observed in some cultivars (Mehlenbacher and Smith, 1991, 2006). Self-incompatibility, which is controlled by a single S-locus with 33 alleles (Mehlenbacher, 2014), requires either that compatible pollinators are planted inside the orchards, or that compatible cultivars are planted close to each other in surrounding orchards. Chilling requirements are different for catkins and pistillate flowers (i.e., the glomerules) and different degrees of dichogamy are frequently observed (Mehlenbacher, 1991). Some cultivars are generally protogynous and other protandrous, but hazelnut dichogamy can also be influenced by local edaphic and climatic conditions (Mehlenbacher, 1991; Solar and Stampar, 2011; Cristofori et al., 2018). Glomerules can grow along the shoots as lateral buds, at the terminal part of each shoot or on catkins peduncles (Germain, 1994). The percentage of glomerules at each position is cultivar-dependent (Hosseinpour et al., 2013), with high-yielding cultivars presenting pistillate flowers associated with catkins (Thompson et al., 1996). This evidence supports the use of pollen concentration as a proxy of the abundance of both male and female flowers. Other factors can contribute to weaken the relationship between released pollen and hazelnut yield, i.e., biological (e.g. pests, diseases) and environmental constraints, (e.g. weather) or agricultural management (e.g. pruning, Pearse et al., 2016).

We aimed here at using pollen data collected in four Chilean hazelnut orchards in the period 2013–2017 to fit statistical models, whose parameters were correlated with yield variability. Our objectives are threefold: i) to select the best suited modelling approach for reproducing seasonal pollen concentrations, ii) to identify the patterns of alternate bearing in the hazelnut yield data and iii) to quantify the relationships between the hazelnut pollen data and yield variability. This work represents a fundamental step towards the realization of an operational hazelnut yield forecasting system, considering aerobiological, agronomic and meteorological as input data.

## 2. Materials and methods

The methodological workflow followed in this study is presented in Fig. 1. Hazelnut airborne pollen concentration and high-resolution yield data were collected in four Chilean hazelnut plantations (Camarico, CAM, Los Niches, LOS, San Sebastian, SAN, Caracas, CAR), considering three main cultivars (Barcelona, Barc, Tonda di Giffoni, TG, Tonda Gentile delle Langhe, TGdL), in the growing seasons 2013/2014–2017/2018. Pollen data were modelled using gaussian and logistic functions, and the parameters of the best-performing model in reproducing seasonal pollen concentration were computed. Hazelnut alternate bearing was characterized by analyzing yield data series, separately for each cultivar and site. Finally, the parameters from both modelled and observed pollen concentrations were correlated with hazelnut yields, to quantify the strength of their relationships and to reveal any geographical or cultivar pattern.

### 2.1. Experimental data collection

Hazelnut airborne pollen concentration data were collected in four hazelnut orchards located in the fertile longitudinal valley be-

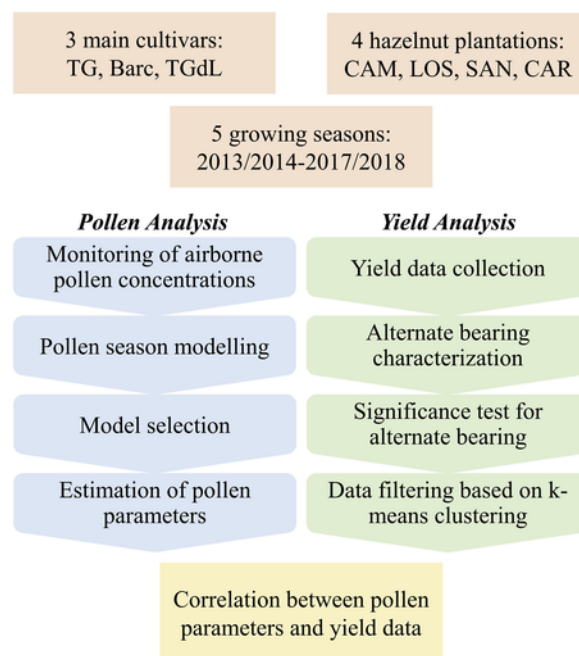


Fig. 1. Methodological workflow followed in this study. Please see the main text for complete explanation.

tween Maule (CAM, LOS SAN) and Araucanía regions in central Chile (CAR, Fig. 2). The hazelnut orchards were planted between the mid-90s and the early 2000s, and their extension ranges from 146 ha LOS to 1461 ha CAM. Each orchard is divided in multiple plots sharing the same irrigation systems, with dripline irrigation as the most frequent method. Soil nutrients were supplemented ranging 90–110N, 30–40P and 50–80K ( $\text{kg ha}^{-1} \text{year}^{-1}$ ), while mechanical and chemical control of suckers and weeds were usually performed 3–4 times a year. Main pests were generally controlled by 2–3 applications a year for insect and 3–4 applications for bacterial infections.

The three hazelnut plantations in the Northern region are characterized by a Mediterranean climate, while CAR has a Marine West Coast climate according to the Köppen–Geiger classification (Sarricolea et al., 2017). The four plantations present a dry and warm summer, with precipitations concentrated mainly during the cool winter season, especially in CAR. Soils are generally sedimentary, with volcanic sediments originating from different geological eras (Casanova et al., 2013). Particularly, in CAR the soil is characterized by a higher organic matter content (local agronomists, personal communication). Different cultivars were planted in each orchard, aiming at increasing pollination capability and at differentiating hazelnut production. Effective pollination is granted by cross-compatibility between plots or by the interposition of compatible cultivars along rows. A multi-stem training system has been employed in the four plantations, with planting density ranging from 666 plants  $\text{ha}^{-1}$  ( $5 \text{ m} \times 3 \text{ m}$ ) to 333 plants  $\text{ha}^{-1}$  ( $6 \text{ m} \times 5 \text{ m}$ ). We focused our analysis on the three main hazelnut varieties cultivated in the four orchards (91.7 % of the yield data), i.e. Barc ( $S_1$   $S_2$ , 35.7 % of the yield data), TGdL ( $S_2$   $S_7$ , 28.3 % of the yield data) and TG ( $S_2$   $S_{23}$ , 27.7 % of the yield data). These cultivars were present in CAM, LOS and SAN, whereas in CAR, TGdL was missing. Hazelnut airborne pollen concentration was monitored by volumetric spore traps (Lanzoni VPPS® Hirst-type sampler), which were placed at canopy level (Fig. 2). The volumetric traps (air suction rates =  $101 \text{ min}^{-1}$ ) contained an adhesive tape (33.3 cm) moved at  $2 \text{ mm h}^{-1}$ , where pollen grains were deposited, and were weekly inspected. Each tape was cut into daily segments and placed on microscope slides with a mounting medium containing basic fuchsin (0.08 % gelatin, 0.44 % glycerin, 0.015 % liquefied phenol, 0.0015 % basic fuchsin in aqueous solution) to selec-



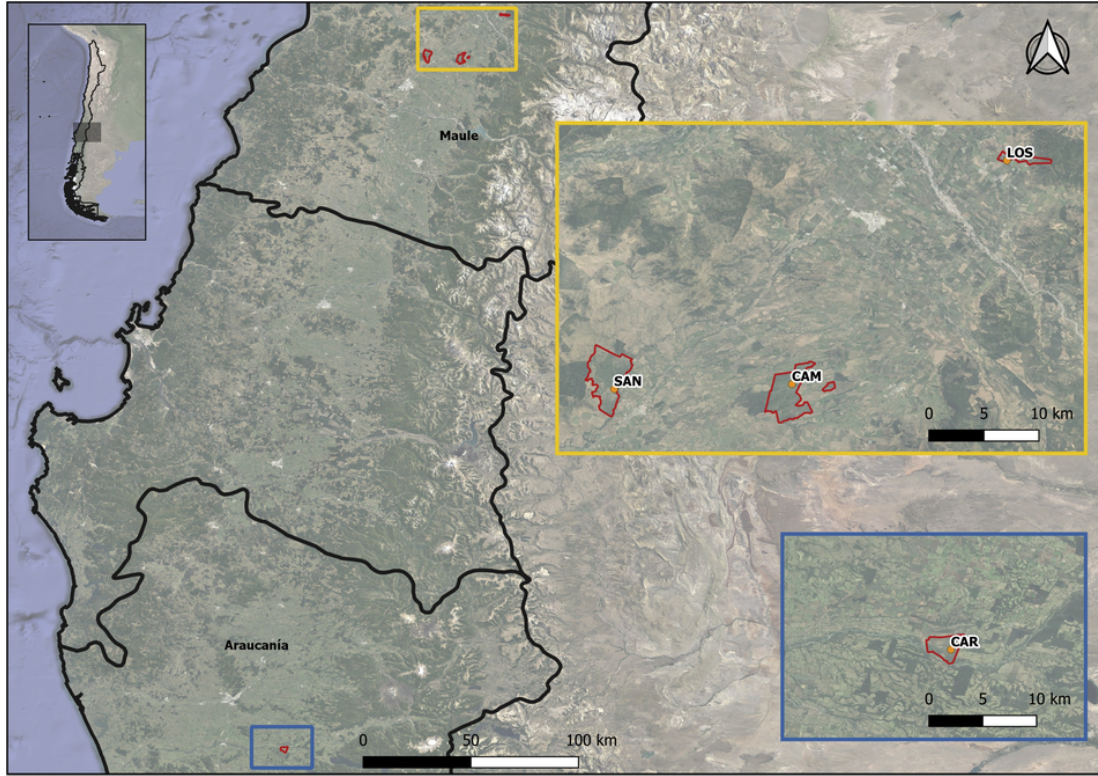


Fig. 2. Geographical distribution of the four Chilean hazelnut plantations and of spore trap locations.

tively stain pollen walls (Ogden et al., 1974). Daily pollen concentrations (pollen  $\text{m}^{-3}$ ) were estimated following the standard protocol used by the Italian Association of Aerobiology (Mandrioli, 2000). Pollen grains coming from different hazelnut cultivars (and species) cannot be distinguished based on morphological differences. Therefore, the measured airborne pollen concentration referred to the flowering of all the cultivars present in each plantation. Hazelnut production (kg) has been collected separately for each plot by mechanized harvesting, and the corresponding yield ( $\text{kg ha}^{-1}$ ) was calculated by dividing the production by the harvested area.

## 2.2. Statistical models of hazelnut airborne pollen concentrations

The data of hazelnut airborne pollen concentrations were fitted by alternative forms of the gaussian (Eq. 1, Eq. 4) and logistic (Eq. 2, Eq. 3) models using the day of the year (DOY) as independent variable (Kasprzyk and Walanus, 2014). This procedure was iteratively applied at each site and growing season, (i) on airborne pollen spectra (APS, pollen grains  $\text{m}^{-3}$ , Eq. 1, Eq. 2), as done by Kasprzyk and Walanus (2014) in modelling atmospheric dynamics of pollen grains and fungal spores, and (ii) on cumulated pollen concentrations (CPC, cumulated pollen grains  $\text{m}^{-3}$ , Eq. 3, Eq. 4), as described by Cunha et al. (2015) for *Vitis* pollen season.

$$P(d) = k \times e^{\left(-\frac{1}{2} \times \frac{(\text{DOY} - \mu)^2}{\sigma^2}\right)} \quad (1)$$

$$P(d) = k \times 4 \times \frac{e^{\left(-\frac{(\text{DOY} - \mu)}{\sigma}\right)}}{\left(1 + e^{\left(\frac{(\text{DOY} - \mu)}{\sigma}\right)}\right)^{-2}} \quad (2)$$

where  $P(d)$  is the daily pollen concentration (pollen grains  $\text{m}^{-3} \text{ d}^{-1}$ );  $k$  is the peak of the pollen dispersal (pollen  $\text{m}^{-3}$ );  $\mu$  is the day of the

year corresponding to the pollen season peak;  $\sigma$  estimates the standard deviation of the distribution, as a measure of pollen season length.

$$cP(d) = \alpha \times \left(1 + e^{-(\beta + \gamma \times \text{DOY})}\right)^{-1} \quad (3)$$

where  $cP(d)$  is the cumulated pollen grains concentration (pollen grains  $\text{m}^{-3}$ ) up to the current DOY;  $\alpha$ , is the distance between the two asymptotes, as an estimate of the total amount of cumulated pollen (cumulated pollen  $\text{m}^{-3}$ );  $\beta$  is a correction factor (unitless);  $\gamma$  represents the rate of increase of pollen dispersal (unitless). Finally, the equivalent gaussian cumulated density function was computed (Eq. 4):

$$cP(d) = k \times \left(1 + \text{erf}\left(\frac{\text{DOY} - \mu}{\sigma \sqrt{2}}\right)\right) \quad (4)$$

where  $cP(d)$  is the cumulated pollen grain concentration (pollen grains  $\text{m}^{-3}$ ) up to the current DOY;  $\mu$  is the DOY at which the cumulated pollen is halved;  $k$  corresponds to the amount of cumulated pollen up to  $\mu$  (cumulated pollen grains  $\text{m}^{-3}$ );  $\sigma$  estimates the model standard deviation, as a measure of pollen season length. The peak of pollen dispersal is not a parameter that could be directly extracted from the logistic or the gaussian models fitted on CPC, therefore the first derivatives of the Eqs 3 and 4 were calculated. The maximum value of the first derivative (i.e., the maximum rate of change) was assumed as the peak of the modelled pollen season (Fig. 3). All the statistical models were fitted (Fig. 3) on pollen data with least square regression using the 'nls tools' package (Baty et al., 2015) in R environment (R Core Team, 2018). The objective function was the Residual Sum of Squares (RSS) and the fitting process was carried out avoiding eventual local minima by the inspection of RSS contours through Beale's criterion (Milliken et al., 1990) as implemented in 'nls tools'. The Bayesian Information Criterion (BIC, Schwartz, 1978) was computed for each model and the one obtaining the minimum BIC was assumed as the best model.

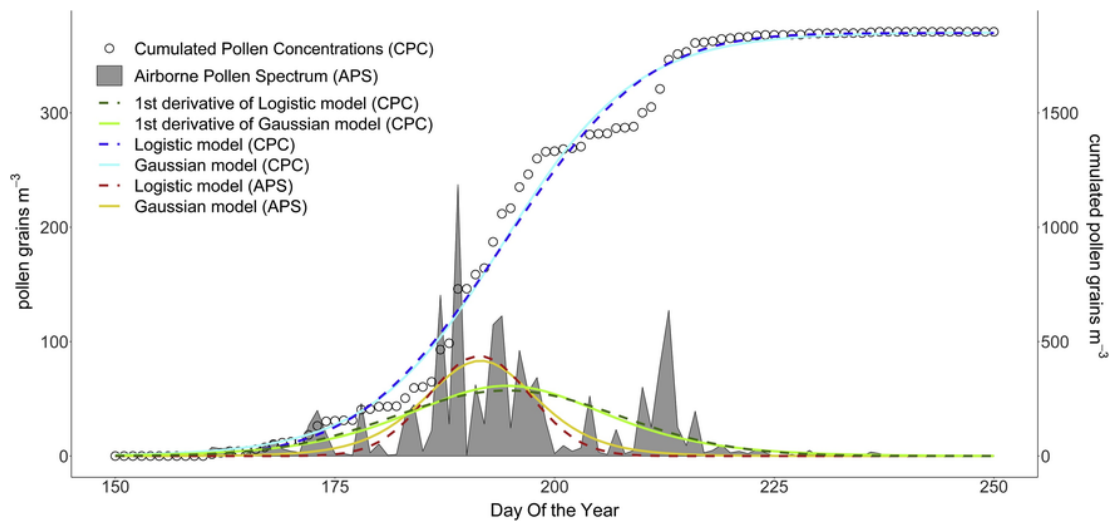


Fig. 3. Logistic and gaussian models fitted on cumulated pollen concentrations and on airborne pollen spectrum together with the estimated first derivative of the logistic model fitted on cumulated pollen concentrations. Sample data of 2016/2017 growing season from San Sebastian (SAN) spore-trap.

### 2.3. Characterization of the alternate bearing in hazelnut plantations

We analyzed continuous hazelnut yield data series in 2013–2018, coming from mature plots, with tree age above 8 years. The presence of alternate bearing on the yield data was tested using the alternate bearing Index ( $I$ ) developed by Hoblyn et al. (1937), which was already adopted for the evaluation of yield variability in different crop species (Rosenstock et al., 2010; Guitton et al., 2012; Sanderson and Treeby, 2014):

$$I = \frac{\sum_{i=2}^n \frac{|y_i - y_{i-1}|}{(y_i + y_{i-1})}}{(n-1)} \quad (5)$$

where  $I$  is the total sum of the absolute difference between two consecutive yields ( $|y_i - y_{i-1}|$ ) divided by their sum ( $y_i + y_{i-1}$ ), averaged by the number of yield data ( $n$ ). According to this formula,  $I = 1$  corresponds to no hazelnut yield in the OFF year (maximum alternate bearing) and  $I = 0$  indicates constant yield (no alternate bearing). A significance test on  $I$  values was then applied to assess their statistical relevance (Huff, 2001): hazelnut yields were resampled with replacement 5000 times and a new index was calculated from each series of resampled values (5000  $I$  values  $\times$  92 plots). The frequency the values of resampled  $I$  exceeded or equalled original  $I$  values corresponds to the probability ( $P$ ) of random alternate bearing in the original data. In other terms, the higher the  $P$ , the lower the likelihood that  $I$  computed from actual yields describes an actual alternate bearing attitude. The distribution of  $P$  was clustered using k-means algorithm in order to select only the plots with an evident biennial bearing (low  $P$  values). The best number of clusters was found using the R 'NbClust' package (Charrad et al., 2015). Significant differences between clusters was tested with Generalized Linear Models (GLMs) and pairwise comparison within cluster means were performed using Tukey's post-hoc test ( $P < 0.05$ ).

### 2.4. Correlations between pollen concentration and hazelnut yield

Observed and modelled pollen concentration data were used to estimate the parameters, which were correlated with yield data. The parameters describing the start and finish of the pollen season, and its middle, length and peak were derived from observed APS data using the R package 'pollen' (Kasprzyk and Walanus, 2014) following the approach described by Andersen (1991). Pollen season start (oSTART, day

of year) and finish (oFINISH, day of year) were defined when the 2.5 % and 97.5 % of total pollen concentration was reached. Pollen season length (oLENGTH, days) included the 95 % of pollen data. The peak (oPEAK, pollen grains  $m^{-3}$ ) and the middle of the pollen season (oMIDDLE, day of year) were considered in the day when the maximum pollen concentration was reached. Finally, the sum of the seasonal pollen concentration was computed (oTOT, pollen grains  $m^{-3}$ ). The parameters estimated using modelled pollen data were computed according to Cunha et al. (2015). The first, second and third curve inflection points corresponded respectively to the modelled start (mSTART, day of year), middle (mMIDDLE, day of year) and finish (mFINISH, day of year) of the pollen season. mSTART and mFINISH were used to calculate pollen season length (mLENGTH, days), while the modelled peak (mPEAK, pollen grains  $m^{-3}$ ) was derived from the model first derivative (Fig. 3). The sum of modelled pollen concentrations was then estimated (mTOT, pollen grains  $m^{-3}$ ). After checking for data normality, the pollen parameters calculated on modelled and observed data were correlated with hazelnut yields using non parametric Spearman's rank correlation. The analysis was distinctly performed on each plantation (i.e. CAM, LOS, SAN and CAR) and cultivar (i.e. Barc, TGdL and TG). The correlation coefficients ( $\rho$ ), the two tailed probability of  $t$  ( $P < 0.05$ ) and the probability of replication ( $P_{rep}$ , Killeen (2005)) for each correlation were computed using the R 'psych' package (Revelle, 2015). The  $P_{rep}$  is the probability that an exact replication study will find the same significance level  $P$ .

## 3. Results

### 3.1. Hazelnut pollen season modelling: model selection and pollen parameter characterization

The pollen dynamics in 2013/2014–2017/2018 growing seasons in the four Chilean plantations were best represented (lowest BIC values) by the logistic function fitted on CPC data (Eq. 3). The first derivative of this model (Eq. 4) was able to correctly reproduce APS in all sites and years (

Fig. 4), providing the best estimate of the peaks of the pollen season.

The models directly fitted on the airborne pollen spectra (Eq. 1, Eq. 2), albeit almost identical in fitting performance, were not capable to reproduce pollen seasons where either a main pollen peak stood out against the remaining spectrum, or when multiple pollen peaks were present (Fig. S1). Both gaussian and logistic models (Eq. 3, Eq. 4)

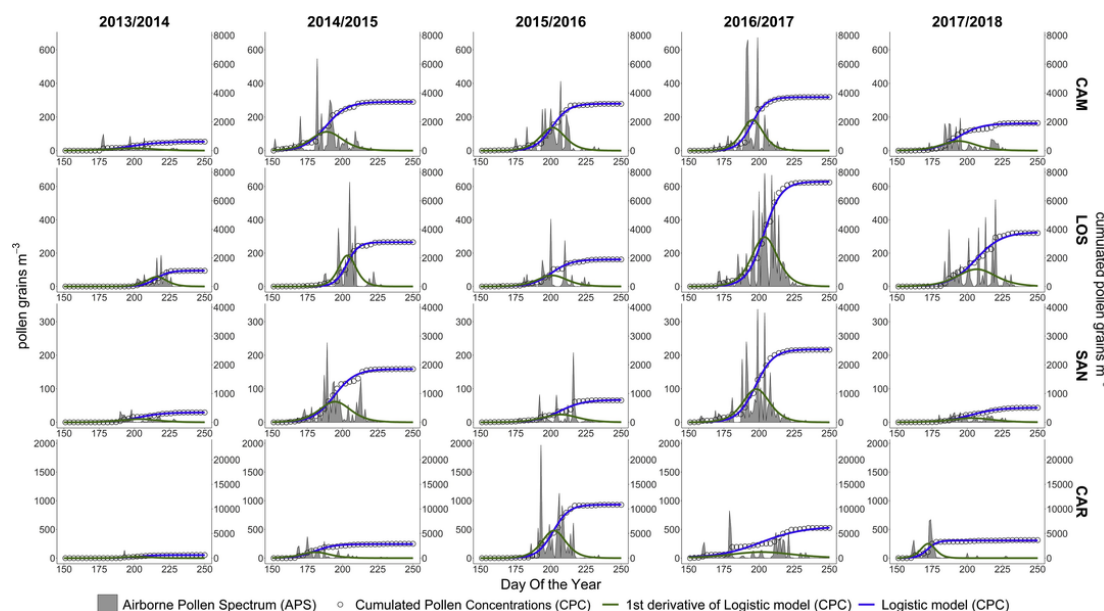


Fig. 4. Logistic models fitted in 2013/2014–2017/2018 growing seasons on cumulated pollen concentrations and first derivative of the fitted models used to approximate airborne pollen spectra. CAM = Camarico, LOS = Los Niches, SAN = San Sebastian and CAR = Caracas.

were able to correctly fit cumulated pollen observations (CPC) (Fig. S1), resulting in similar performances (Table S2). Nevertheless, the lowest BIC was obtained in most cases by the logistic model (Table S3), which was retained in the following analyses.

Table 1 shows the comparison between the measured pollen concentration data and the values of the parameters of the logistic model. The modelled peak and the pollen season length had the highest coefficient of variation (CV) compared to observed values (average values for peak = 81 % vs 68 %; average values for length = 35 % vs 19 %). The differences between average modelled and observed cumulated concentrations (TOT) were very low (Mean values = 15, CV = 0 %), therefore only modelled values were retained for correlation analyses. On average, the CAR plantation, situated in the Araucanía region where winter precipitations are more abundant with cooler temperatures, led to the highest modelled pollen amount (4870 grains  $m^{-3}$ ). The three sites LOS, SAN and CAR had strong CVs for modelled pollen peaks during the studied years (71 %, 93 % and 93 %, respectively). The two northern sites of LOS and SAN shared the same pattern of pollen variability, while CAR showed the opposite behavior (

Fig. 4). CAM plantation presented a quite stable pollen production (Fig. 4) and a lower CV for modelled peak (66 %) compared to the other sites (Table 1).

Table 1

Coefficient of variation (CV) and mean values for modelled (“m” prefix) and observed (“o” prefix) pollen parameters in the four Chilean hazelnut plantations. The last column holds overall average values. (CAM = Camarico, LOS = Los Niches, SAN = San Sebastian and CAR = Caracas). Please see section 2.4 for the explanation of the acronyms.

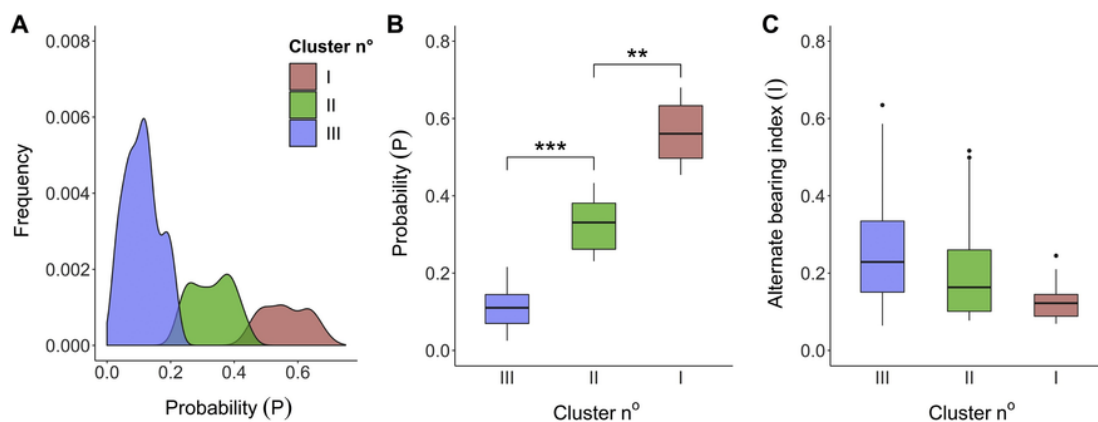
	CAM		LOS		SAN		CAR		Overall	
	CV (%)	Mean	CV (%)	Mean	CV (%)	Mean	CV (%)	Mean	CV (%)	Mean
mPEAK	66	100	71	142	93	41	93	193	81	119
oPEAK	62	385	41	484	78	170	89	795	68	458
mTOT	50	2574	70	3451	79	1204	81	4870	70	3025
oTOT	50	2578	70	3426	79	1199	80	4837	70	3010
mSTART	4	178	4	191	2	182	7	174	4	181
oSTART	5	170	6	183	5	175	8	173	6	175
mMIDDLE	2	195	3	206	2	201	8	192	4	198
oMIDDLE	6	192	4	210	5	199	5	183	5	196
mFINISH	3	212	3	220	3	220	10	209	5	215
oFINISH	3	222	2	225	3	225	4	228	3	225
mLENGTH	29	35	32	29	17	38	61	34	35	34
oLENGTH	11	53	25	42	9	50	30	54	19	50

### 3.2. Analysis of hazelnut alternate bearing

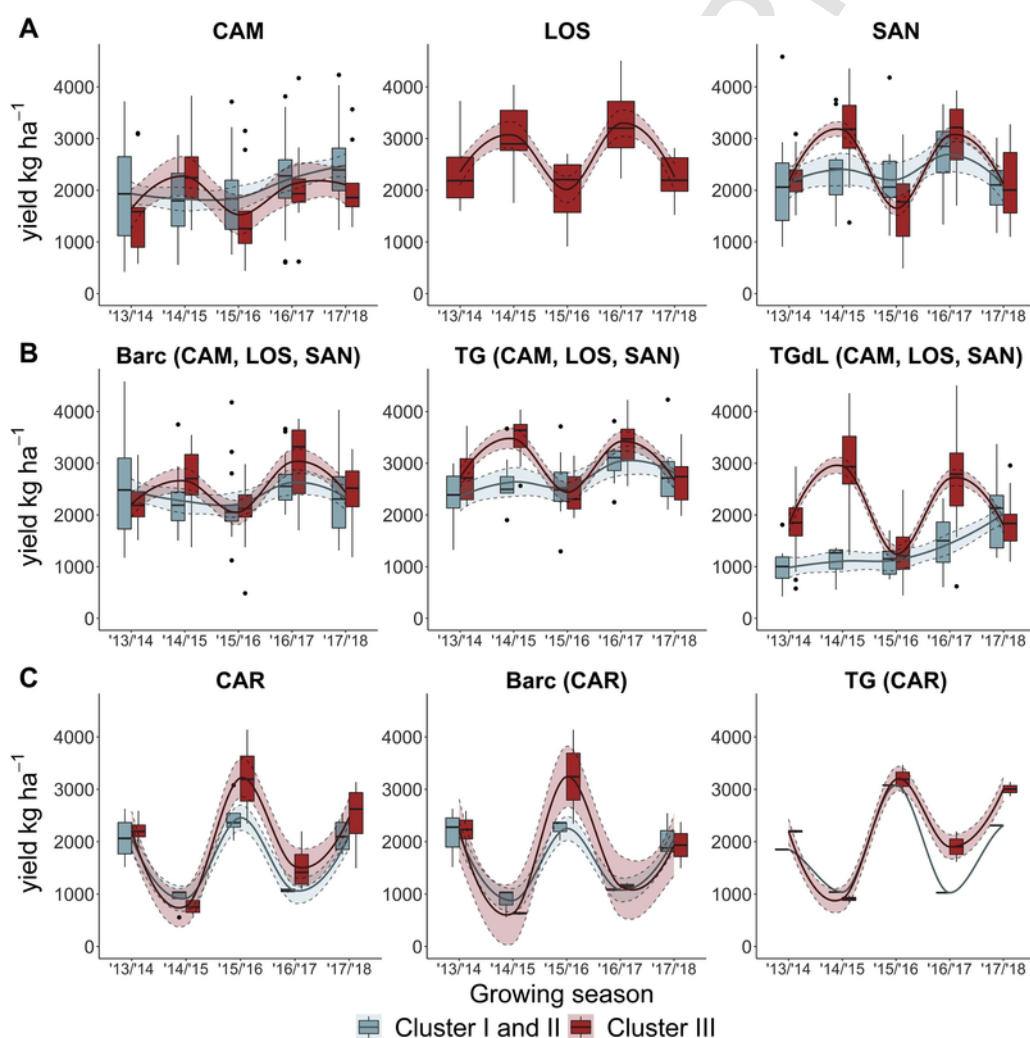
The estimation of the *I* index (Eq. 5, Hoblyn et al., 1937) computed on consecutive yield data series and the clustering of the probabilities (*P*) resulting from the resampling with replacement allowed identifying three clusters corresponding to distinct patterns of alternate bearing (Fig. 5).

Cluster I (average  $P = 0.56 \pm sd = 0.07$ ) and II (average  $P = 0.33 \pm sd = 0.06$ ) included respectively 16.5 % and 24.8 % of the yield data, and corresponded to plots with a limited alternate bearing. Conversely, cluster III grouped 58.8 % of the yield data, and corresponded to a stronger alternate bearing attitude (average  $P = 0.1 \pm sd = 0.05$ ). The average plant age increased from cluster I to III and the values between clusters were significantly different ( $I = 12.8$  years,  $II = 14.1$  years,  $III = 15.3$  years,  $P < 0.05$ ), suggesting that older plots are associated with a more pronounced alternate bearing. The distributions of hazelnut yields according to the three clusters are shown in Fig. 6 as boxplots. Although only 8.8 % of the plots showed significant values ( $P < 0.05$ ), the clustering algorithm allowed identifying the plots characterized by an alternate bearing (cluster III), and the ones characterized by more constant yields (cluster I and II). CAM pre-





**Fig. 5.** (A) The distribution of probabilities of error (P) in rejecting the hypothesis that the alternate bearing in original data is a non-random phenomenon were clustered using k-means. (B) Differences between cluster average fRoNRI as assessed using Generalized Linear Model (GLMs) and Tukey's post-hoc test ( $P < 0.05$ ). (C) Alternate bearing values (I) for the identified clusters.



**Fig. 6.** Hazelnut yield data in 2013–2018 grouped according to k-means filtering and by site (A) and by cultivar (B) in the northern plantations. (C) Effect of k-mean filtering of hazelnut yields in the southern plantation and relative cultivars. The sites were split to underline the opposite alternate bearing. Data from clusters I and II were combined. Fitted is a Loess smoothing with 0.80 confidence interval. (CAM = Camarico, LOS = Los Niche, SAN = San Sebastian, CAR = Caracas, Barc = Barcelona, TG = Tonda di Giffoni, TGDl = Tonda Gentile delle Langhe).

sented an overall regular hazelnut production (I = 47 %, II = 25 %, III = 28 % of yield data), while a stronger alternate bearing characterized LOS (III = 100 % of yield data), SAN (II = 26 %, III = 74 % of yield data) and CAR (II = 50 %, III = 50 % of yield data). CAM, LOS and SAN, sit-

uated in the northern region, shared the same alternate pattern, with 2013/2014, 2015/2016 and 2017/2018 as OFF growing seasons. Conversely, the southern orchard CAR and the two main cultivars (Barc and TG) presented the opposite biennial bearing (Fig. 6). TGDl (I =

9 %, II = 32 %, III = 59 % of yield data) showed the strongest alternate bearing, followed by TG (I = 2.2 %, II = 7.7 %, III = 14 % of yield data) and Barc (I = 6 %, II = 22 %, III = 40 % of yield data), regardless of the plantation.

### 3.3. Correlation between pollen parameters and filtered yields

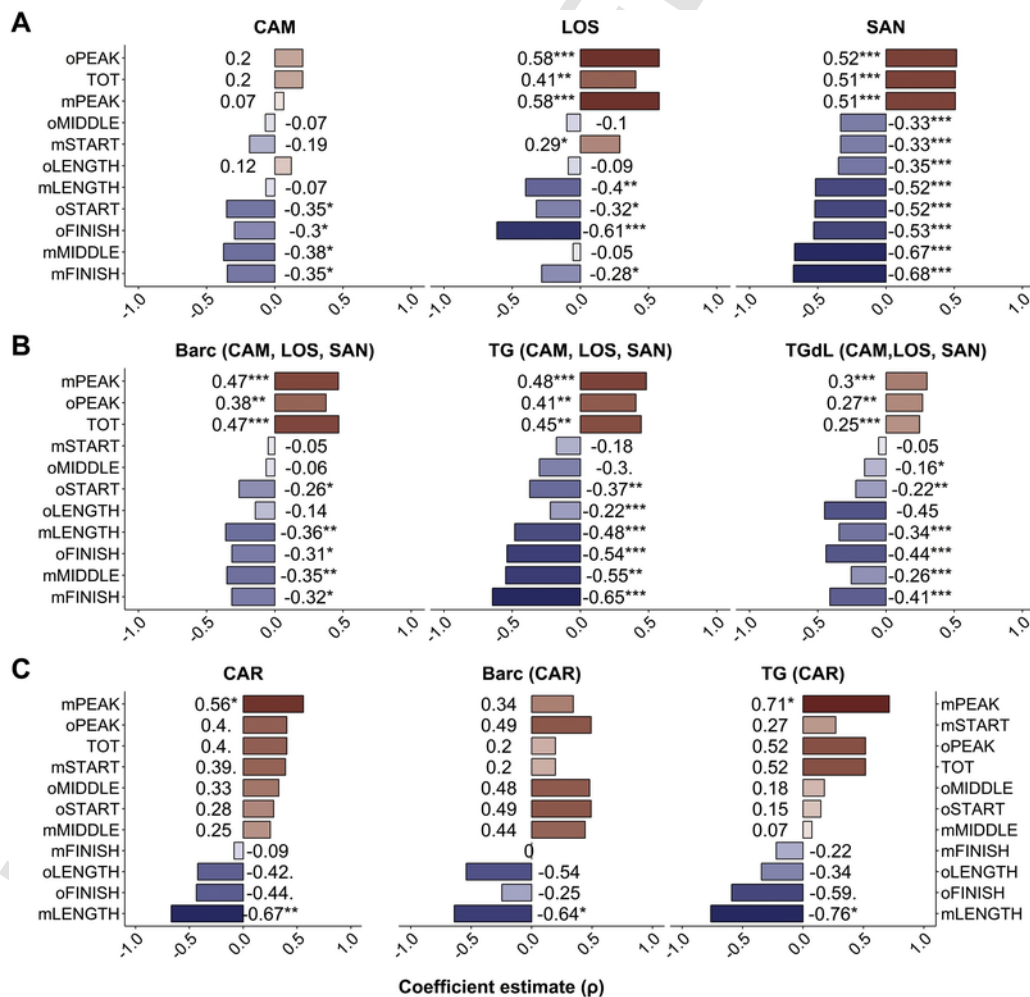
The analysis of the correlations between the parameters of the statistical models fitted on the pollen concentration data and the corresponding yield variability revealed significant correlations, considering the data grouped by site and cultivar. The yield-pollen correlation was stronger when only data in cluster III were retained, rather than using the whole yield dataset (Fig. 7, and Fig. S2). The average absolute Spearman correlation coefficients ( $\rho$ ) were higher both considering the data divided by plantation ( $\rho = 0.36$  vs  $\rho = 0.29$ ) and by cultivar ( $\rho = 0.35$  vs  $\rho = 0.27$ ). The parameters describing the pollen emission intensity (TOT, PEAK) were always positively associated to yields, whereas parameters describing pollen season shape and position (START, FINISH, MIDDLE and LENGTH) negatively correlated with hazelnut yields (Fig. 7). Focusing on the correlations obtained with yield data in cluster III, mPEAK obtained the highest average correlation coefficient ( $\rho = 0.45$ ) compared to TOT ( $\rho = 0.38$ ) and oPEAK ( $\rho = 0.40$ ). The average power of the association was lower in CAM ( $\rho = 0.21$ ) and higher in LOS ( $\rho = 0.34$ ), SAN ( $\rho = 0.50$ ) and CAR ( $\rho = 0.39$ ). The yield alternate bearing more closely followed

pollen season trends in Barc ( $\rho = 0.33$ ) and TG ( $\rho = 0.41$ ) than in TGDl ( $\rho = 0.29$ ). Detailed results with significance levels (P), replication probabilities ( $P_{rep}$ ) and sample sizes (n) are presented in the supplementary Table S2 (by site) and Table S3 (by cultivar).

## 4. Discussion

### 4.1. Unravelling the relationships between hazelnut airborne pollen and yield

Airborne pollen concentrations collected in the four Chilean hazelnut plantations showed a clear inter-annual variability in the considered growing seasons (Fig. 3). A strong alternate bearing also characterized hazelnut yields grouped by sites and cultivars (Fig. 6). Moreover, yield and pollen production were significantly related (Fig. 7). Following the “wind-pollination” hypothesis, mast seeding, i.e., the simultaneous and variable fruit production between years of perennial plants (Pearse et al., 2016), would be more beneficial for wind-pollinated, self-incompatible species as a mean to guarantee cross-pollination, and to produce higher quality seeds while decreasing in-breeding depression (Kelly et al., 2001). The results of our study agree with this hypothesis, as the correlations between yields and pollen peaks were positive, while the ones with pollen season duration were negative (Fig. 7), indicating that a short and intense male flowering is beneficial to fruit production. A positive correlation between the annual sum



**Fig. 7.** Spearman correlations ( $\rho$ ) between yields in cluster three and parameters calculated from modelled pollen data (“m”) and parameters from non-modelled pollen data (“o”). (A) Orchards situated in Maule, (B) Cultivars planted in the orchards located in Maule, (C) Orchard situated in Araucanía and relative cultivars. Significance levels: \*\*\* < 0.001, \*\* < 0.01, \* < 0.05, < 0.1 (CAM = Camarico, LOS = Los Niches, SAN = San Sebastian, CAR = Caracas, Barc = Barcelona, TG = Tonda di Giffoni, TGDl = Tonda Gentile delle Langhe).



of daily pollen concentrations and acorn yields was also observed in *Quercus ilex* (García-Mozo et al., 2007). Conversely, our results tend to exclude detrimental effects of excessive pollen concentrations on female flower viability, as observed in *Juglans* spp., where pollen data were negatively associated to fruit-set (Prentović et al. 2014). We also rule out yield limitations due to excessive amounts of self or incompatible pollen on stigmas that would have contributed in undermining the relationship between pollen and crop production.

Spore-traps and hazelnut harvest also allowed detecting an opposite alternate bearing in the two Chilean regions where plantations are located. Further investigations should try to disentangle whether local precipitation patterns or extreme weather events could influence local hazelnut biennial bearing and the relationships between airborne pollen and crop yield.

Finally, our results estimate the dependence of pollen-yield correlation on alternate bearing power (Fig. 7 and Fig. S2), making pollen concentrations a putative parameter for the modelling of yield variability. Airborne pollen has been used as a predictor to increase the performance of many yield forecasting systems. A high percentage of yield variability was explained when the total amount of released pollen was used to forecast wine production in Portugal (Cunha et al., 2016). Olive annual pollen concentrations were also strongly associated to yields in the Mediterranean basin both at a national and regional level (Oteros et al., 2014).

#### 4.2. Assumptions in modelling airborne pollen concentrations

The approach followed by Kasprzyk and Walanus (2014) who directly applied logistic and gaussian models to fungal spores and pollen emission spectra, failed to correctly represent APS in our study when the pollen season was characterized by strong multimodality or in presence of an evident maximum peak (Fig. S1, supplementary material). In those cases, multimodal distributions would have been more appropriate although adding complexity in the process of parameter selection. Conversely, cumulated pollen concentrations were suitably approximated by both gaussian and logistic distributions (Fig. 4). Our data support the use of the logistic model as the best approach to describe cumulated pollen concentrations, as found by other authors (Ribeiro et al., 2007; Cunha et al., 2015).

The calculation of curve inflection points (Cunha et al., 2015) allowed excluding the noise associated with pollen season start and finish and resulted in higher CVs of the pollen season length (average CV = 35 %). Moreover, the peak of the pollen season calculated from the model first derivative, was found to maximize CVs in all the considered sites (average CV = 81 %), making it a better proxy for flowering inter-annual variability.

Several studies proposed alternative indicators to describe the main features of the pollen season, and used them as input in yield forecasting models. Some authors relied on actual pollen data, selecting cumulated pollen concentrations as the main predictor (García-Mozo et al., 2007; Oteros et al., 2013). Nevertheless, other researchers have underlined the limits of considering the sum of cumulated pollen concentrations as estimator of flowering intensity, as it is usually affected by locally re-suspended pollen (Fornaciari et al., 2002; Orlandi et al., 2005). A refined approach consisted in applying statistical models on pollen concentration data, and adopting the derived parameters to explain yield variability, leading to a higher accuracy in yield forecasting (Cunha et al., 2016). Our results support this evidence, showing that a proper modelling approach bring benefits in characterizing pollen seasons with variable duration and intensity.

#### 4.3. A methodology to identify hazelnut alternate bearing

The k-means clustering approach used in this study allowed identifying 58.8 % of hazelnut yield data series with a strong alternate bearing, although only 8% were significant. The computation of the alternate bearing index and the adoption of the resampling technique (Huff, 2001) have been used for the estimation of alternate bearing in pistachio, where 58 % of the analysed plants showed significant alternate bearing (Rosenstock et al., 2010). During the analysis of de-trended orange yield data, an evidence of alternate bearing was found in three out of twelve cultivars (Sanderson and Treeby, 2014). Huff (2001) demonstrated that increasing yield trends in the tree juvenile phase affect the statistical significance of the alternate bearing index. Smith et al. (2004) also showed that  $I$  (Eq. 5) increases and  $P$  decreases in mandarin, demonstrating that alternate bearing is intrinsically related with tree age.

Although only yield data from plants at least eight years old were considered in our analysis, an increasing yield trend was still visible in clusters II and I, where alternate bearing was smaller. Moreover, the plant age increased from cluster I to III, meaning that some of the analysed plants were still in an active growth phase. An alternate bearing attitude clearly emerged when plants were around 15 years old (cluster III). TGD L cultivar was particularly affected by alternate bearing as confirmed by the expertise of hazelnut growers.

Overall, our data confirmed the value of using the alternate bearing index coupled with resampling (Hoblyn et al., 1937; Huff, 2001). Recently, the consecutive disparity index ( $D$ ) was introduced as a measure of temporal variability and has been applied to evaluate mast seeding of European forest trees (Fernández-Martínez et al., 2018). Further analysis of hazelnut alternate bearing should compare the performance of different indices over longer time series and on different cultivars to better characterize this phenomenon.

## 5. Conclusions

Investigating the relationships between airborne pollen concentration and yield is fundamental in nut crops, where fruits number depends on the flowering intensity, in turn subjected to a marked alternate bearing. The methodology proposed here allowed providing evidences of the bearing attitude in Chilean hazelnut plantations, and led to quantifying the strength of the association between pollen concentration data and yield variability considering the main cultivars. The consideration of aerobiological parameters emerges as a crucial aspect to refine hazelnut yield prediction, as integrated in a forecasting system considering the impact of agronomic management and of weather conditions during the growing season.

## Uncited reference

CRediT authorship contribution statement

**Lorenzo Ascari:** Investigation, Formal analysis, Writing - original draft, Visualization. **Consolata Siniscalco:** Project administration, Resources. **Giovanna Palestini:** Investigation. **María José Lisperguer:** Investigation. **Eloy Suarez Huerta:** Funding acquisition, Supervision. **Tommaso De Gregorio:** Funding acquisition, Supervision. **Simone Bregaglio:** Conceptualization, Methodology, Writing - review & editing.

## Declaration of Competing Interest

None.

## Acknowledgements

The research was carried out with the cooperation and contribution of the Ferrero Hazelnut Company (HCo); particular appreciation goes to Laura Giustarini (HCo) for assisting in data visualization; an important contribution for gathering field data was provided by Agrichile Staff.

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.eja.2020.126036>.

## References

- Andersen, T.B., 1991. A model to predict the beginning of the pollen season. *Grana* 30, 269–275.
- Baty, F., Ritz, C., Charles, S., Brutsche, M., Flandrois, J.-P., Delignette-Muller, M.L., 2015. The R package nlstools: a toolbox for nonlinear regression. *J. Stat. Softw.* 66, 1–21.
- Boccacci, P., Aramini, M., Valentini, N., et al., 2013. Molecular and morphological diversity of on-farm hazelnut (*Corylus avellana* L.) landraces from southern Europe and their role in the origin and diffusion of cultivated germplasm. *Tree Genet. Genomes* 9, 1465–1480.
- Casanova, M., Salazar, O., Seguel, O., Luzio, W., 2013. *The Soils of Chile*. Springer Berlin Heidelberg.
- Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A., 2015. NbClust : An R package for determining the relevant number of clusters in a data set. *J. Stat. Softw.* 61, 1–36.
- Cristofori, V., Pica, A.L., Silvestri, C., Bizzarri, S., 2018. Phenology and yield evaluation of hazelnut cultivars in Latium region. *Acta Hort.* 1226, 123–130.
- Cunha, M., Ribeiro, H., Costa, P., Abreu, I., 2015. A comparative study of vineyard phenology and pollen metrics extracted from airborne pollen time series. *Aerobiologia* 31, 45–56.
- Cunha, M., Ribeiro, H., Abreu, I., 2016. Pollen-based predictive modelling of wine production: application to an arid region. *Eur. J. Agron.* 73, 42–54.
- FAOSTAT, 2020. FAO Statistics, Food and Agriculture Organization of the United Nations. In: <http://www.fao.org/faostat/en/>.
- Fernández-Martínez, M., Vicca, S., Janssens, I.A., Carnicer, J., Martín-Vide, J., Peñuelas, J., 2018. The consecutive disparity index, D: a measure of temporal variability in ecological studies. *Ecosphere* 9.
- Fornaciari, M., Pieroni, L., Orlandi, F., Romano, B., 2002. A new approach to consider the pollen variable in forecasting yield models. *Econ. Bot.* 56, 66–72.
- Galán, C., García-Mozo, H., Vázquez, L., Ruiz, L., Díaz De La Guardia, C., Domínguez-Vilches, E., 2008. Modeling olive crop yield in Andalusia, Spain. *Agron. J.* 100, 98–104.
- García-Mozo, H., Gómez-Casero, M.T., Domínguez, E., Galán, C., 2007. Influence of pollen emission and weather-related factors on variations in holm-oak (*Quercus ilex* subsp. *ballota*) acorn production. *Environ. Exp. Bot.* 61, 35–40.
- Germain, E., 1994. The reproduction of hazelnut (*Corylus avellana* L.): a review. *Acta Hort.* 351, 195–210.
- Guitton, B., Kelner, J.J., Velasco, R., Gardiner, S.E., Chagné, D., Costes, E., 2012. Genetic control of biennial bearing in apple. *J. Exp. Bot.* 63, 131–149.
- Hampson, C.R., Azarenko, A.N., 1993. Pollen-stigma interactions following compatible and incompatible pollinations in hazelnut. *J. Am. Soc. Hort. Sci.* 118, 814–819.
- Heslop-Harrison, Y., Heslop-Harrison, J.S., Heslop-Harrison, J., 1986. Germination of *Corylus avellana* L. (hazel) pollen: hydration and the function of the oncus. *Acta Bot. Neerl.* 35, 265–284.
- Hoblyn, T.N., Grubb, N.H., Painter, A.C., Wates, B.L., 1937. Studies in biennial bearing.—I. *J. Pomol. Hort. Sci.* 14, 39–76.
- Hosseinpour, A., Seifi, E., Molnar, T.J., 2013. Position and density of pistillate inflorescences of some hazelnut cultivars grown in Iran. *J. Agric. Sci.* 5, 125–131.
- Huff, A., 2001. A significance test for biennial bearing using data resampling. *J. Hort. Sci. Biotechnol.* 76, 534–535.
- Kasprzyk, I., Walanus, A., 2014. Gamma, Gaussian and logistic distribution models for airborne pollen grains and fungal spore season dynamics. *Aerobiologia* 30, 369–383.
- Kelly, D., Hart, D.E., Allen, R.B., 2001. Evaluating the wind pollination benefits of mast seeding. *Ecology* 82, 117–126.
- Killeen, P.R., 2005. An alternative to null-hypothesis significance tests. *Psychol. Sci.* 16, 345–353.
- Mandrioli, P., 2000. Method for sampling and counting of airborne pollen and fungal spores. Institute of Atmospheric and Oceanic Sciences, National Research Council, Bologna.
- Mehlenbacher, S.A., 1991. Chilling requirements of hazelnut cultivars. *Sci. Hort.* 47, 271–282.
- Mehlenbacher, S.A., 2014. Geographic distribution of incompatibility alleles in cultivars and selections of European hazelnut. *J. Am. Soc. Hort. Sci.* 139, 191–212.
- Mehlenbacher, S.A., Smith, D.C., 1991. Partial self-compatibility in “Tombul” and “Montebello” hazelnuts. *Euphytica* 56, 231–236.
- Mehlenbacher, S.A., Smith, D.C., 2006. Self-compatible seedlings of the cutleaf hazelnut. *HortScience* 41, 482–483.
- Milliken, G.A., Bates, D.M., Watts, D.G., 1990. Nonlinear regression analysis and its applications. *Technometrics* 32, 219.
- Molnar, T.J., 2011. *Corylus*. In: Kole, C. (Ed.), *Wild Crop Relatives: Genomic and Breeding Resources*. Springer, Berlin Heidelberg, pp. 15–48.
- Nuts and Dried Fruits Statistical Yearbook, 2020. International Nut and Dried Fruit Council, In: <https://www.nutfruit.org/industry/statistics>.
- Ogden, E.C., New York State Museum and Science Service, U.S. Atomic Energy Commission, 1974. *Manual for Sampling Airborne Pollen*. Hafner Press.
- Orlandi, F., Romano, B., Fornaciari, M., 2005. Effective pollination period estimation in olive (*Olea europaea* L.): a pollen monitoring application. *Sci. Hort.* 105, 313–318.
- Orlandi, F., Carlo, S., Tommaso, B., Luigia, R., Bruno, R., Marco, F., 2010. Yield modelling in a Mediterranean species utilizing cause-effect relationships between temperature forcing and biological processes. *Sci. Hort.* 123, 412–417.
- Oteros, J., García-Mozo, H., Hervás, C., Galán, C., 2013. Biometeorological and autoregressive indices for predicting olive pollen intensity. *Int. J. Biometeorol.* 57, 307–316.
- Oteros, J., Orlandi, F., García-Mozo, H., et al., 2014. Better prediction of Mediterranean olive production using pollen-based models. *Agron. Sustain. Dev.* 34, 685–694.
- Pearse, I.S., Koenig, W.D., Kelly, D., 2016. Mechanisms of mast seeding: resources, weather, cues, and selection. *New Phytol.* 212, 546–562.
- Prentović, M., Radišić, P., Smith, M., Šikoparija, B., 2014. Predicting walnut (*Juglans* spp.) crop yield using meteorological and airborne pollen data. *Ann. Appl. Biol.* 165, 249–259.
- R Core Team, 2018. *R: a Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Revelle, W., 2015. *Psych: Procedures for Personality and Psychological Research*.
- Ribeiro, H., Cunha, M., Abreu, I., 2007. Definition of main pollen season using a logistic model. *Ann. Agric. Environ. Med.* 14, 259–264.
- Ribeiro, H., Abreu, I., Cunha, M., 2017. Olive crop-yield forecasting based on airborne pollen in a region where the olive groves acreage and crop system changed drastically. *Aerobiologia* 33, 473–480.
- Rosenstock, T.S., Rosa, U.A., Plant, R.E., Brown, P.H., 2010. A reevaluation of alternate bearing in pistachio. *Sci. Hort.* 124, 149–152.
- Sanderson, G.P., Treeby, M.T., 2014. Fruit yield of common orange varieties in south-east Australia: initial evaluation versus long term performance. *Sci. Hort.* 180, 111–117.
- Sarricolea, P., Herrera-Ossandon, M., Meseguer-Ruiz, Ó., 2017. Climatic regionalisation of continental Chile. *J. Maps* 13, 66–73.
- Schwartz, G., 1978. Estimating the dimension of a model. *Ann. Stat.* 6, 461–464.
- Smith, H.M., Samach, A., 2013. Constraints to obtaining consistent annual yields in perennial tree crops. I: heavy fruit load dominates over vegetative growth. *Plant Sci.* 207, 158–167.
- Smith, M.W., Shaw, R.G., Chapman, J.C., et al., 2004. Long-term performance of “Ellendale” mandarin on seven commercial rootstocks in sub-tropical Australia. *Sci. Hort.* 102, 75–89.
- Solar, A., Stampar, F., 2011. Characterisation of selected hazelnut cultivars: phenology, growing and yielding capacity, market quality and nutraceutical value. *J. Sci. Food Agric.* 91, 1205–1212.
- Thompson, M., Lagerstedt, H., Mehlenbacher, S., 1996. *Hazelnuts*. Fruit Breeding. John Wiley & Sons, Inc., New York, 125–184.